DOCUMENT RESUME

ED 312 298 TM 014 070

AUTHOR Campbell, Kathleen T.
TITLE Dangers in Using Analysis of Covariance Procedures.
PUB DATE Nov 89
NOTE 14p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (17th, Louisville, KY, November 9-11, 1988).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Covariance; Analysis of Variance; *Multivariate Analysis; Regression (Statistics); Researchers; *Research Problems; Statistical Analysis; Theory Practice Relationship

ABSTRACT

Problems associated with the use of analysis of covariance (ANCOVA) as a statistical control technique are explained. Three problems relate to the use of "OVA" methods (analysis of variance, analysis of covariance, multivariate analysis of variance, and multivariate analysis of covariance) in general. These are: (1) the wasting of information when intervally scaled independent variables are converted to the nominal level; (2) the distortion of distribution shapes of and relationships among the non-interval predictor variables; and (3) the reduction of power against Type II error. Three other problems are associated with the use of ANCOVA as a statistical control technique: the need for very reliable measurement of the control variables; the regard many researchers have for ANCOVA as an almost magical technique for equalizing dissimilar groups; and the fact that researchers frequently disregard the critical homogeneity of regression assumption. If the regression equations of the groups are not reasonably similar, the single regression equation calculated by ignoring group membership will result in underadjustment for the experimental group. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312298

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

KATHLEEN T. CAMPBELL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DANGERS IN USING ANALYSIS OF COVARIANCE PROCEDURES

Kathleen T. Campbell

University of New Orleans

Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY, November 9, 1988.

TM014070

BEST COPY AVAILABLE

ABSTRACT

Researchers have historically used analysis of covariance (ANCOVA) to make statistical adjustments in intact groups, as in analyzing the effectiveness of programs such as Head Start, in order to minimize the differences which exist between experimental and control groups at the start of an experiment. The paper intends to explain the problems associated with the use of ANCOVA as a statistical control technique. Three problems relate to the use of OVAs in general: (1) the wasting of information when intervally scaled independent variables are converted to the nominal level; (2) the distortion of distribution shapes of and relationships among non-interval predictor variables; and (3) the reduction of power against Type II error.

There are three other problems associated with ANCOVA as a statistical control technique. The first involves the often overlooked but crucial assumption of very reliable measurement of the control variables. The second involves the regard that many researchers hold toward ANCOVA as an almost magical technique for equalizing dissimilar groups. The primary difficulty with ANCOVA, however, involves the critical homogeneity of regression assumption which is often disregarded by researchers. If the regression equations of groups are not reasonably similar, then the single regression equation calculated by ignoring group membership will result in an underadjustment for the experimental group.

Dangers in Using Analysis of Covariance Procedures

A statistical control procedure, analysis of covariance (ANCOVA), is used by researchers in quasi-experimental or ex post facto designs to make groups equivalent when random selection or assignment is not possible or desirable. The procedure entails making an adjustment on the dependent variable, using one or more covariates, in a regression adjustment that completely ignores group membership. The adjustment is expected to minimize the initial difference between the groups.

There are some inherent problems with the use of ANCOVA, however. The first three problems relate to the problems of the use of OVAs (ANOVA, ANCOVA, MANOVA, MANCOVA) in general. First, since "OVA methods require that all independent variables be nominally scaled" (Thompson, 1986a, p. 918), and since most independent variables are higher than nominally scaled (e.g., intervally scaled), this results in the wasting of much information. As Thompson (1981) notes, "When we reduce interval level of scale data to the nominal level of scale we are doing nothing less than thoughtlessly throwing away information which we previously went to some trouble to collect" (p. 8).

The second problem associated with OVAs is that these "methods distort the distribution shapes of and relationships among non-interval predictor variables" (Thompson, 1986b, p. 18). Furthermore, most researchers employing these designs

use balanced designs of "exactly equal numbers of subjects per cell" (Thompson, 1986a, p. 918). This is done so that "all sums of squares for effects when cumulated will exactly equal the total sums of squares for the dependent variable" (Thompson, 1986a, p. 918). Although this allows for "computational simplicity" (Cohen, 1968, p. 440), computational simplicity is not so necessary in the age of widespread use of computers.

The third problem associated with OVAs is that these "methods tend to reduce power against Type II error by reducing the reliability levels of variables that were originally higher than nominally scaled" (Thompson, 1986a, p. 19). Thus, by reducing intervally scaled variables to the nominal level, OVAs both lessen reliablilty and raise the likelihood of a Type II error, i.e., reduce the probability of achieving statistically significant results.

Since an ANCOVA is actually an ANOVA procedure performed on the residualized dependent variable scores (Y minus YHAT), the three problems associated with OVAs in general apply equally to ANCOVA. However, there are additional problems associated with ANCOVA, in particular, as a statistical control technique. As noted earlier, ANCOVA is sometimes used to adjust findings when random assignment or random selection was not possible or "when the quantitative researcher believes that random selection or random assignment or design selection have failed to create groups that were equivalent at the start of the experiment or quasi-

experiment" (Thompson, 1986b, p. 19).

The first problem associated with ANCOVA and statistical controls in general is "that they assume very reliable measurement of the control variables" (Thompson, 1986b, p. 20). As Nunnally (1975, p. 10) notes "[m]easurement reliablilty becomes crucial... in employing statistical partialling operations, as in the analysis of covariance or in the use of partial correlational analysis." Many researchers, however, do not even report the measurement error of their variables, and may inappropriately make statistical corrections using unreliable covariates that make random adjustments.

A second problem associated with the use of ANCOVA is that many researchers who were not able to obtain random assignment or selection of their subjects seem to regard the statistical control as an almost magical method for making unlike groups equivalent. Unfortunately, ANCOVA is not a panacea for equalizing dissimilar groups.

The main difficulty with using statistical controls in order to make groups equivalent involves the homogeneity of regression assumption. As Thompson (1986b, p. 22) notes:

> This assumption is necessary because the statistical control procedures are implemented by adjusting the dependent variable to the extent that the covariate and the dependent variable are correlated <u>when group membership information is ignored</u>.

An intuitive explanation of ANCOVA is given by Huck, Cormier, and Bounds (1974). They give an example of statistically adjusting for differences between two groups who have been pretested, and who then received two different teaching methods, lecture and discussion, and were then posttested with a final exam which was identical to the pretest. On the pretest, or covariate, the lecture (control) group received a higher mean score, 14.5, than did the discussion (experimental) group, which received a 9.5, whereas on the posttest, or dependent variable, the mean score of 34.8 of the lecture group was only 2.7 points higher than the mean score of 32.1 of the discussion group. An analysis of covariance was used to adjust for initial differences of the two groups. As Huck, Cormier, and Bounds explain (p. 134):

> In a nonscientific manner, our researcher could make this adjustment by first averaging the two pretest means to find out the mean score for all subjects, disregarding group membership, on the pretest. This would result in an overall pretest mean of 12.0. Since the lecture group had a pretest mean that was 2 1/2 points higher than the overall average, this group's final exam mean must be reduced by 2 1/2 points to account for the fact that the students in this group began the course with a head start. Thus, the adjusted final exam mean for the lecture group becomes equal to 34.8

minus 2.5, or 32.3. On the other hand, the discussion group had a pretest mean that was 2 1/2 points below the overall average; therefore, this group's final exam mean must be increased by 2 1/2 points to account for the fact that the students in this group began the course with a disadvantage. Thus, the adjusted final exam mean for the discussion group becomes equal to 32.1 plus 2.5, or 34.6.

Although this explanation is a severe oversimplification of the actual procedure, the logic holds true conceptually *if and only if* the researcher meets a critical analytic assumption. In order to do this legitimately, the two groups must meet the homogeneity of regression assumption; i.e., *regression equations computed separately for the groups must be reasonably similar to each other*. This is because ANCOVA makes the statistical adjustment using a regression equation derived by ignoring group membership, and this adjustment will therefore only be legitimate if the equations for the groups are similar enough that the use of one common equation is reasonable.

An example of two groups whose regression slopes violate the homogeneity of regression assumption can be seen on the graph in Figure 1. Using the hypothetical data set in z form from Thompson (1986b, p. 24, Table 1), the slopes for groups A and B were calculated and plotted separately, as represented by the dotted lines. One common regression

equation was then calculated, ignoring group membership, and plotted, as represented by the solid line. In an ANCOVA procedure, the statistical adjustment would be made on this "line of best fit," despite the fact that the two regression slopes which it purports to represent are quite dissimilar. The statistical adjustment would result in an underadjustment for the experimental group A. While ANCOVA can control for the initial head start of group B, the procedure can not control for the superior rate at which group B continues to learn. Thus, the statistical adjustment can control for the initial difference between groups but not for the continuing difference of different learning rates.

Insert Figure 1 about here

Researchers have historically used ANCOVA to make statistical adjustments in intact groups, as in analyzing the effectiveness of compensatory education programs such as Head Start. The Head Start program was given to all eligible students. Because of the disadvantaged background of these students in their early formative years, there was a wide gap between their knowledge base and that of average students. Not only was there a gap in the knowledge base, however, but the average students also learned at a much faster rate. While Head Start was expected to remediate the disadvantaged students, it was never expected to be a miracle cure which would not only bridge the gap of knowledge bases but would

also increase their learning rate to the equivalence of the average students. Yet, in applying an analysis of covariance as a statistical control, that is exactly what the researchers were implying. While Head Start may help to bridge the gap somewhat on the knowledge bases, unless it also serves to increase drastically the learning rate of the experimental group, the program will appear to be ineffective, or worse (Campbell & Erlebacher,1975). Analysis of covariance, then, is not useful unless groups' learning slopes (i.e., regression equations) are fairly equivalent in the first place, in which case a statistical control is probably not needed.

Campbell and Erlebacher (1975) present a simulated example to illustrate how ANCOVA can bias results when the homogeneity of regression assumption is not met. Evaluations of compensatory education programs, such as Head Start, are usually quasi-experimental or ex post facto since the treatment is usually given to all eligible children, minimizing the possibility of obtaining random selection or assignment. But the untreated population, or control group, is usually more able than the experimental group.

> In such a situation the usual procedures of selection, adjustment, and analysis produce systematic biases in the direction of making the compensatory program look deleterious... These biases of analysis occur both where pretest scores are available and in ex post facto studies.

> It seems reasonably certain that this methodological error occurred in the Westinghouse-Ohio University study... and it probably has occurred in others purporting to show no effects or harmful effects from Head Start programs (Campbell & Erlebacher, 1975, p. 597).

Campbell and Erlebacher note that, although there have been a few isolated warnings about other statistical control procedures, such as matching, the warning message is newer for ANCOVA. The stated purpose of their essay was to illustrate with a detailed example why these statistical control procedures lead to biased and distorted results. They reported that, "Nevertheless we will be able to show that even in the present clear-cut case of no treatment effects, the common quasi-experimental analysis techniques [including ANCOVA] will result in serious biases" (Campbell & Erlebacher, 1975, p. 598). Using a simulated data set with absolutely no treatment effect, they showed that the underadjustment of the experimental group through the use of ANCOVA made the experiment look worse than ineffective:

> The underadjustment by the analysis of covariance has commonly been overlooked, and the resulting bias makes the statistical criticisms of the Westinghouse-Ohio University study by Smith and Bissell (1970) seem trivial in comparison.... We can confidently conclude that, had the Head Start programs actually produced no effects whatsoever,

the mode of analysis used in the Westinghouse-Ohio University study would have made them look worse than useless, actually harmful. (Campbell & Erlebacher, 1975, p. 608)

As Thompson (1986b, p. 23) has noted about the analysis of covariance:

> The statistical control procedure assumes that the relationship between the two variables is the same in both groups, i.e., since correlation is a measure of the slope of the regression line for the two variables, that children who are eligible for and receive compensatory interventions learn at the same rate as children who are not eligible for the intervention. If statistical control is needed because two groups are not equivalent, but the homogeneity of regression assumption is not met, its use often leads to biased results.

As Campbell and Erlebacher (1975) explain:

> The deep-rooted seat of the bias is probably the unexplicit trust that, although the assumptions of a given statistic are technically not met, the effects of these departures will be unsystematic. The reverse is, in fact, true. The more one needs the "controls" and "adjustments" which these statistics seem to offer, the more biased are their outcomes. (Campbell & Erlebacher, 1975, p. 613).

References

Campbell, D. T., & Erlebacher, A. (1975). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In M. Guttentag & E. L. Struening (Eds.), Handbook of evaluation research: Vol. 1. (pp. 597-617). Beverly Hills: SAGE.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Huck, S., Cormier, W., & Bounds, G. (1974). Reading statistics and research. New York: Harper & Row.

Nunnally, J. C. (1975). Psychometric theory--25 years ago and now. Educational Researcher, 4(10), 7-14, 19-20.

Thompson, B. (1981, November). The problem of OVAism. Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington, KY. [Document No. 03980, National Auxillary Publication Service, P. O. Box 3513, Grand Central Station, New York, NY 11017]

Thompson, B. (1986a). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1986b). The place of qualitative research in contemporary social science: The importance of post-paradigmatic thought. Paper presented at the annual meeting (session #21.38) of the American Educational Research Association, San Francisco, April 18, 1986.

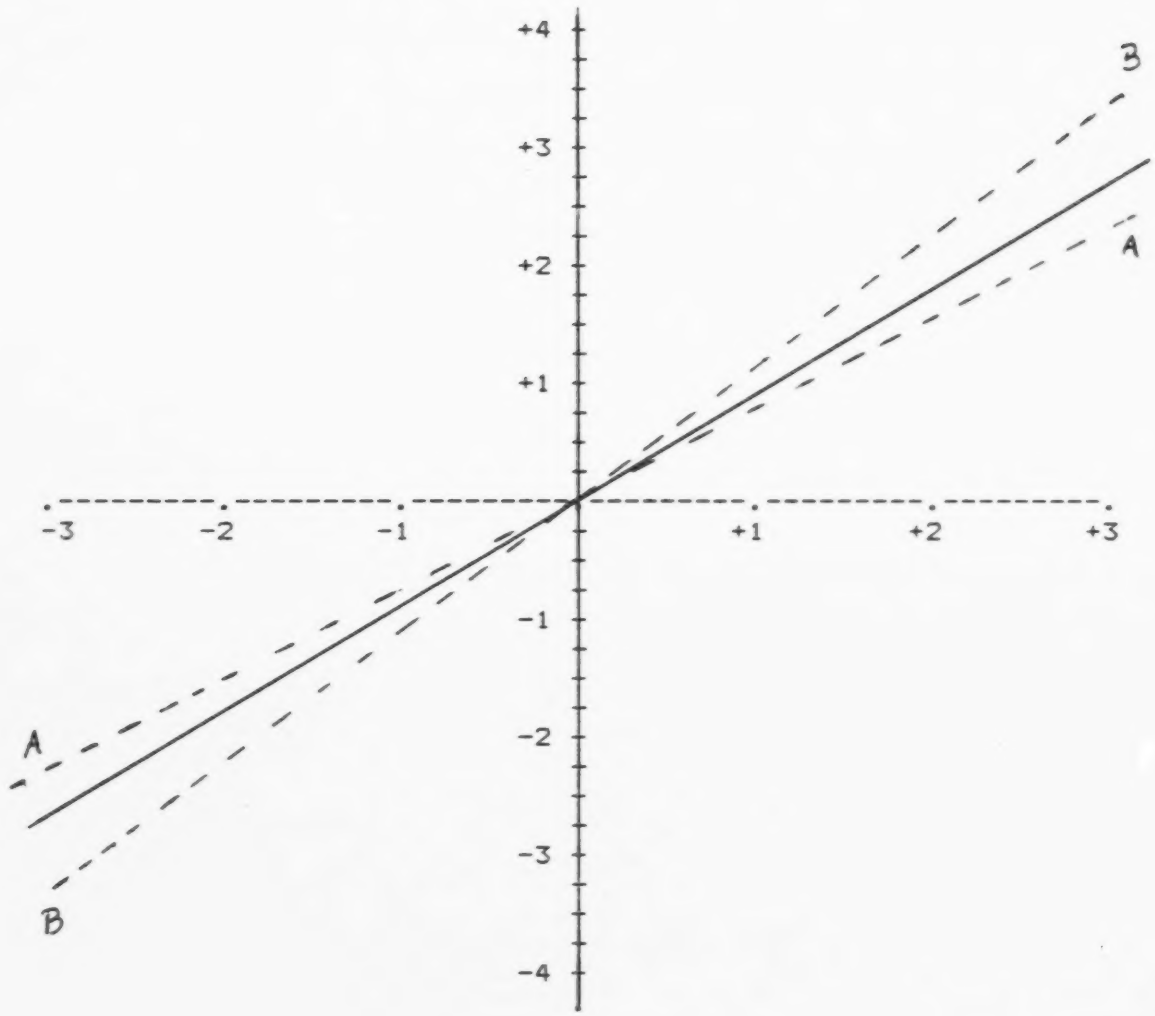

Figure 1. Example of two slopes which violate the homogeneity of regression assumption. (The dotted lines represent two groups with different regression slopes, while the solid line represents the single regression equation calculated by ignoring group membership.)